

<https://helda.helsinki.fi>

Pan-Genomic Analysis of Clostridium botulinum Group II (Non-Proteolytic C. botulinum) Associated with Foodborne Botulism and Isolated from the Environment

Brunt, Jason

2020-05

Brunt , J , van Vliet , A H M , Stringer , S C , Carter , A T , Lindström , M & Peck , M W 2020
, ' Pan-Genomic Analysis of Clostridium botulinum Group II (Non-Proteolytic C. botulinum)
Associated with Foodborne Botulism and Isolated from the Environment ' , Toxins , vol. 12 ,
no. 5 , 306 . <https://doi.org/10.3390/toxins12050306>

<http://hdl.handle.net/10138/317891>
<https://doi.org/10.3390/toxins12050306>

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.




This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Article

Pan-Genomic Analysis of *Clostridium botulinum* Group II (Non-Proteolytic *C. botulinum*) Associated with Foodborne Botulism and Isolated from the Environment

Jason Brunt ^{1,2,*} , Arnoud H. M. van Vliet ³ , Sandra C. Stringer ², Andrew T. Carter ²,
Miia Lindström ⁴ and Michael W. Peck ^{2,*} 

¹ Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK

² Gut Health and Food Safety, Quadram Institute, Norwich Research Park, Norwich NR4 7UQ, UK; sandra.stringer@quadram.ac.uk (S.C.S.); andrewcarter594@btinternet.com (A.T.C.)

³ School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford GU2 7AL, UK; a.vanvliet@surrey.ac.uk

⁴ Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, 00014 Helsinki, Finland; miia.lindstrom@helsinki.fi

* Correspondence: jb2271@cam.ac.uk (J.B.); mike.peck@quadram.ac.uk (M.W.P.)

Received: 22 March 2020; Accepted: 7 May 2020; Published: 8 May 2020



Abstract: The neurotoxin formed by *Clostridium botulinum* Group II is a major cause of foodborne botulism, a deadly intoxication. This study aims to understand the genetic diversity and spread of *C. botulinum* Group II strains and their neurotoxin genes. A comparative genomic study has been conducted with 208 highly diverse *C. botulinum* Group II strains (180 newly sequenced strains isolated from 16 countries over 80 years, 28 sequences from Genbank). Strains possessed a single type B, E, or F neurotoxin gene or were closely related strains with no neurotoxin gene. Botulinum neurotoxin subtype variants (including novel variants) with a unique amino acid sequence were identified. Core genome single-nucleotide polymorphism (SNP) analysis identified two major lineages—one with type E strains, and the second dominated by subtype B4 strains with subtype F6 strains. This study revealed novel details of population structure/diversity and established relationships between whole-genome lineage, botulinum neurotoxin subtype variant, association with foodborne botulism, epidemiology, and geographical source. Additionally, the genome sequences represent a valuable resource for the research community (e.g., understanding evolution of *C. botulinum* and its neurotoxin genes, dissecting key aspects of *C. botulinum* Group II biology). This may contribute to improved risk assessments and the prevention of foodborne botulism.

Keywords: *Clostridium botulinum*; botulism; foodborne; neurotoxin; non-proteolytic

Key Contribution: Core genome single-nucleotide polymorphism (SNP) analysis of 208 highly diverse *Clostridium botulinum* Group II strains has revealed novel details of population structure/diversity and established relationships between whole-genome lineage, botulinum neurotoxin subtype variant, association with foodborne botulism, epidemiology, and geographical source.

1. Introduction

Clostridium botulinum is a polyphyletic taxon classified solely by the ability to form the deadly botulinum neurotoxin. It is separated into four groups (*C. botulinum* Groups I to IV) according to genotypic and phenotypic properties, and the four groups are sufficiently distinct as to be considered

separate species [1–8]. The botulinum neurotoxin is also formed by some strains of *Clostridium baratii* and *Clostridium butyricum*, and a complete putative botulinum neurotoxin-encoding gene has been discovered in *Enterococcus*, with related putative genes in other non-clostridia [3,7,9–12]. Botulinum neurotoxins are the most potent toxins known and act in the cytoplasm of nerve cells, where they selectively cleave proteins involved in neurotransmitter release, resulting in a deadly flaccid paralysis called botulism [13–15]. For more than one hundred years, botulinum neurotoxins have been separated into serotypes using specific antisera to neutralise neurotoxin in animal tests [16,17]. More recently, sequencing of neurotoxin genes and the derivation of the amino acid sequence has led to the identification of subtypes within botulinum neurotoxin serotypes. A standardized numerical notation has been used for botulinum neurotoxin subtype nomenclature, and it has been proposed that a new subtype should diverge by at least 2.6% in amino acid sequence from an existing subtype [16]. More than 40 botulinum neurotoxin subtypes have been described, each comprising variants with a unique coding (and predicted amino acid) sequence. Interestingly, evidence is becoming available that botulinum neurotoxin subtypes display functional variation and possess distinct toxicological properties [18–22].

C. botulinum Group II isolates form a single type B, E, or F neurotoxin. Closely related non-toxic strains are also frequently described and are known as non-toxic (or non-neurotoxic) *C. botulinum* Group II. The botulinum neurotoxin is associated with accessory proteins in various complexes. Genes encoding the neurotoxin and accessory proteins are co-located in one of two conserved clusters (*ha* cluster or *orfX* cluster). Type B neurotoxin genes are located on a small multicopy plasmid (47–63 kb) in a *ha* cluster that comprises genes encoding the neurotoxin, non-toxic-non-haemagglutinin (NTNH), three haemagglutinins, and a positive regulator (*botR*). Type E and F neurotoxin genes are generally located on the chromosome (but for a small fraction of strains, the type E neurotoxin gene is present on a single-copy plasmid) in an *orfX* cluster that comprises genes encoding the neurotoxin, NTNH, and four open reading frames of unknown function [3,16,17,23–31].

Strains of *C. botulinum* Group II are widely distributed in the environment and may enter the food chain [32–35], potentially leading to foodborne botulism. Foodborne botulism is a severe and deadly intoxication caused by consuming food containing as little as 50 ng of botulinum neurotoxin [6,36]. *C. botulinum* Group II strains forming type B or type E neurotoxin are most frequently associated with foodborne botulism. Isolation of the first strain of *C. botulinum* was reported by Emile van Ermengem in 1897 following an outbreak of foodborne botulism involving salted ham in Belgium [37]. From the description of the bacterium, this was a strain of *C. botulinum* Group II that probably formed botulinum neurotoxin type B [16,17]. Unfortunately, this strain has been lost.

C. botulinum Group II (also known as non-proteolytic *C. botulinum* or psychrotrophic *C. botulinum*) is a saccharolytic bacterium with a minimum growth temperature of 3 °C that forms spores of moderate heat resistance. Foodborne botulism outbreaks have been associated with temperature abuse of products intended to be stored chilled, including fish and meat, and continued extreme vigilance is necessary to ensure the safe production of mildly processed, vacuum/modified atmosphere-packed and other chilled foods [3,4,23,24,38–41].

Genomic analysis of *C. botulinum* Group II indicates that it can be separated into two (or three) major lineages depending on the strains tested and criterion applied [3,5,42–49]. The majority of strains forming type E neurotoxin are located in one lineage that is distantly separated from the other lineage that is dominated by strains forming subtype B4 neurotoxin and contains strains forming subtype F6 neurotoxin. Strains forming type E neurotoxin are frequently isolated from fish and the arctic/subarctic environment, but have now also been found in the southern hemisphere [6,50,51]. Strains forming subtype B4 neurotoxin have been previously isolated from European terrestrial environments (including pigs) and from marine (notably, north America) environments [23,51].

At the start of the present study, there were only 28 genomes available in Genbank for *C. botulinum* Group II. The purpose of the present study is to provide insights into the genetic diversity and spread of strains of *C. botulinum* Group II and their neurotoxin genes. To achieve this, the genomes of

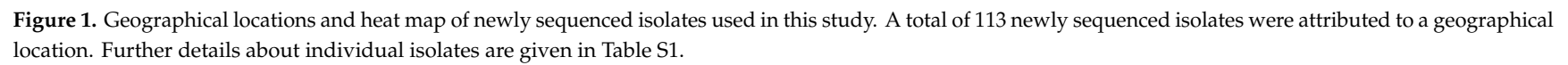
180 additional, highly diverse strains of *C. botulinum* Group II were sequenced. The 208 strains studied were isolated from 16 countries and four continents over a period of over 80 years. This is the most comprehensive comparative genomic study of *C. botulinum* Group II undertaken to date. Additionally, the total of 208 genome sequences represent a valuable resource for the research community—for example, contributing to understanding the evolution of *C. botulinum* Group II and its neurotoxin genes, to the dissection of key aspects of the biology of *C. botulinum* Group II important for food chain transmission, such as spore formation, spore germination [52,53], psychrotrophy [54], neurotoxin formation [55–57], and in the targeting of PCR assays to detect and distinguish strains. Such information will contribute to more comprehensive microbiological risk assessments, the safe development of novel “fresh-like” foods, and the prevention of foodborne botulism.

2. Results and Discussion

2.1. *Botulinum Neurotoxins and Their Encoding Genes*

This study analysed the genomes of 208 strains of *C. botulinum* Group II. This comprised 180 newly sequenced genomes, and 28 genomes from public sources. The newly sequenced genomes considerably increase the number of genome sequences publicly available for *C. botulinum* Group II. The sequenced strains were from a wide geographical area (16 countries over four continents; Figure 1), with the first strain isolated in 1934. Thirty-seven strains were associated with foodborne botulism outbreaks, 94 strains were non-clinical isolates not associated with foodborne botulism outbreaks (i.e., isolated from the environment), and the status of the remaining strains is unknown (Table S1). The strains possessed a single neurotoxin gene that encoded neurotoxin of subtype B4, E1, E2, E3, E6 or F6 or there was no neurotoxin gene. No genes were identified that might encode previously undescribed botulinum neurotoxin serotypes or subtypes. Variants, with a unique coding sequence (and predicted amino acid sequence), were identified for each neurotoxin subtype.

Eight subtypes of type B neurotoxin (B1 to B8) have been described that differ by a maximum of 7.1% amino acid residues [15]. Botulinum neurotoxin subtype B4 is exclusively formed by strains of *C. botulinum* Group II, and a previous survey of 16 examples of this neurotoxin revealed amino acid differences of 1.6% [3,16]. The genome of 63 *C. botulinum* Group II strains examined in the present study contained a subtype B4 neurotoxin gene within a typical *ha* neurotoxin gene cluster arrangement. No other genes encoding entire or partial botulinum neurotoxins were detected in these genomes. From analysis of the B4 botulinum neurotoxin-encoding genes, three distinct variants of this subtype were identified (Figure 2). These three variants have been previously described for a smaller number of strains [3,5,16,23], and it is notable that no new variants were identified in this larger study, suggesting that future subtype B4 neurotoxins may also fall within the same arrangement. Although information is not available for all strains, there is evidence of potential epidemiological and geographic distribution bias. The largest number of strains were in the upper subtype B4 variant group (Figure 2). Where known, these strains were all isolated in Europe, and either associated with foodborne botulism or were isolated from the environment (non-clinical isolates). The foodborne botulism outbreaks involved the consumption of meat in Iceland, Italy and the U.K. [23,58]. The U.K. outbreak was associated with the consumption of home-prepared pork that had been home slaughtered, bottled and kept at ambient temperature for several months in Poland before being brought to the U.K. Following treatment with botulinum antitoxin, the patient made a full recovery [58]. A single subtype B4 variant was recently found in all seven strains isolated in France and associated with six foodborne botulism outbreaks involving meat [43]. All strains in the centre subtype B4 variant group (Figure 2) were from the environment (non-clinical isolates) and were principally from north America. The seven strains in the lower subtype B4 variant group were all isolated in north America, six following outbreaks of foodborne botulism involving fish, and the seventh from Pacific Ocean sediments (Figure 2).



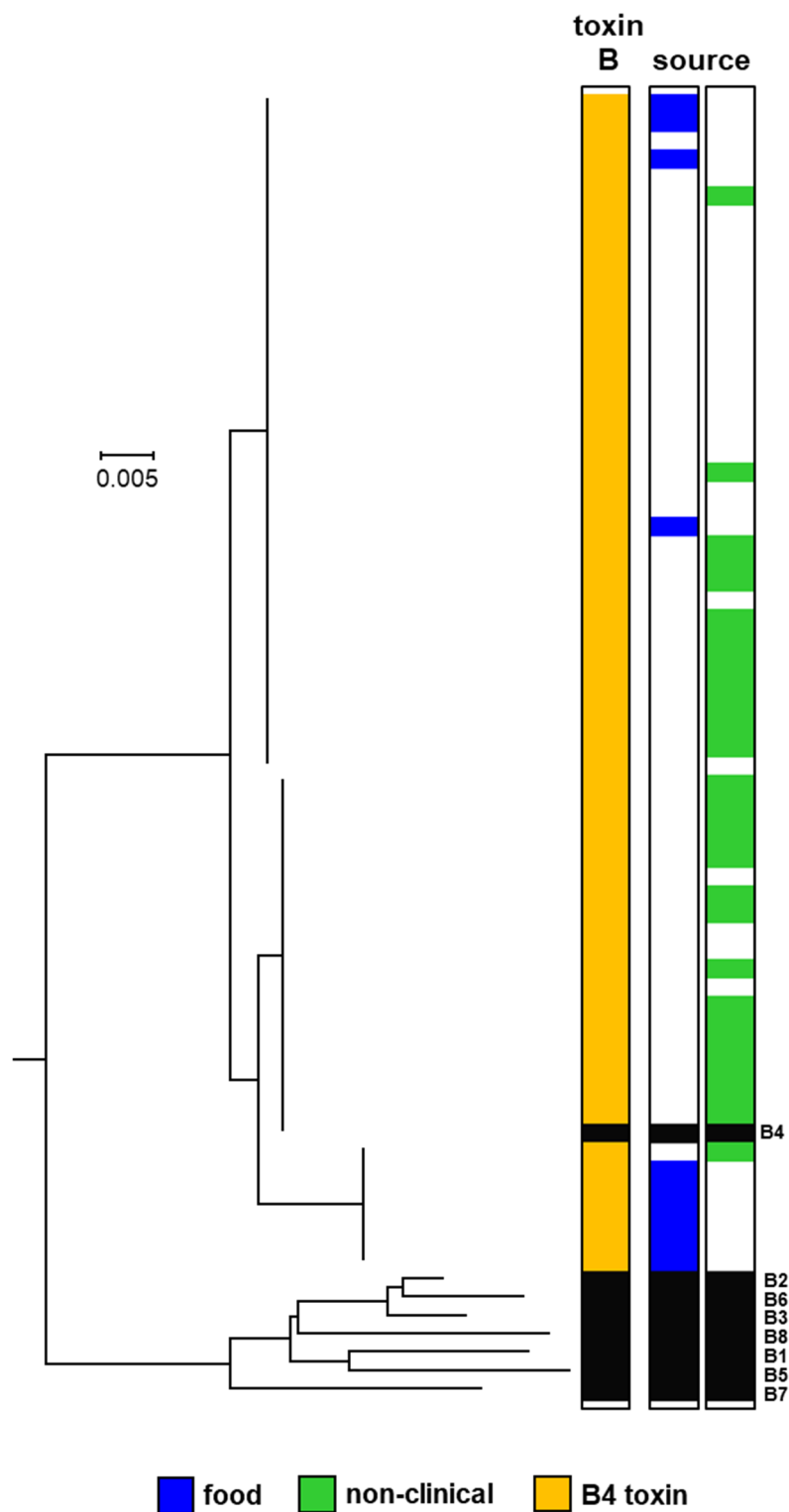


Figure 2. Phylogeny of subtypes of botulinum neurotoxin type B. Further details about individual sequences and isolates are given in Table S2, where they appear in the same order as in this figure. The sources of the isolates are given in Table S1. The toxin protein sequences were aligned with the Muscle module of MEGA7 [59] algorithm and the phylogenetic tree was generated using the Neighbour-Joining method. Three subtype B4 neurotoxin variants are present in *C. botulinum* Group II. Scale bar represents the number of amino acid substitutions per site. A total of 71 amino acid sequences were analysed. White blocks represent absence of information regarding source. Black blocks represent reference neurotoxins [16].

There are currently 12 recognised subtypes of botulinum neurotoxin type E (E1 to E12). These subtypes are generally closely related (amino acid difference of 0.9–5.9%), although subtypes E9 and E12 are more diverse [16]. The genomes of 98 strains contained a type E neurotoxin gene, all belonging to known subtypes [16]. The genome of 58 strains contained a subtype E1 botulinum neurotoxin gene, three strains contained a subtype E2 botulinum neurotoxin gene, 30 strains a subtype E3 gene, and seven strains a subtype E6 gene. This is a valuable addition to the existing sequence data that was heavily focused on subtype E3 [16]. All type E neurotoxin genes were located within a typical *orfX* neurotoxin gene cluster.

Fifty-eight strains possessed a gene encoding subtype E1 botulinum neurotoxin (Figure 3). These strains appear widespread, and the strains examined in the present study have been isolated over a period of more than 70 years from northern Europe (Finland, Sweden, Denmark and Norway), central Europe (Austria and Hungary), north America (Canada and USA), Africa (Egypt), and Asia (Japan), with many associated with foodborne botulism (Table S1). Several variants of subtype E1 botulinum neurotoxin were identified (Figure 3), as noted previously [16,42]. The sequence for subtype E1 neurotoxin derived from whole-genome sequences for two isolates of strain Beluga from our collection (CB0187, CB0603) and one from the USA (CB1085) were identical to each other and also that derived from whole-genome sequences of other strains (Figure 3), but different on eight positions out of 1252 to the E1 toxin in accession number CAA43999, which is commonly used as the standard E1 toxin. In future comparisons, there may be merit in using the subtype E1 neurotoxin sequence in genome accession ACSC01 (CB1085) as the new standard. Interestingly, the present study revealed five amino acid sequences (strains Saroma, SAR, 1304E, IFR 18/112, IFR 18/132) that form two new closely related variants of subtype E1 (highlighted in red box in Figure 3). Two of the strains (Saroma and 1304E) are non-clinical isolates and were probably isolated from lake sludge in Hokkaido Island, Japan, before 1970. Additionally, “SAR” may be an abbreviation for Saroma, indicating a similar origin. Amino acid alignment revealed that four of these new subtype E1 sequences (strains; Saroma, SAR, IFR 18/112, IFR 18/132) shared 98.6% identity with the subtype E1 toxin sequence in genome accession ACSC01 (CB1085), while the toxin of strain 1304E shared 99.0% identity with subtype E1 toxin sequence at the amino acid level (and they also shared 97.9% and 98.3% identity with the E2 neurotoxin, respectively). Guidelines indicate that a 2.6% amino acid difference should be used to separate botulinum neurotoxin subtypes [16]. However, although botulinum neurotoxin subtypes E1, E2 and E3 are distinct, the amino acid differences between these subtypes is less than 2.6%. Thus, if the 2.6% amino acid difference guideline to separate botulinum neurotoxin subtypes was applied, then these would be considered a single subtype [16]. The new subtype E1 variant sequences described here are 1.0% to 1.4% different from subtype E1, and therefore do not meet the above-described criterion for a new botulinum neurotoxin subtype and have therefore been designated as subtype E1. Botulinum neurotoxins consist of two chains, forming an N-terminal light chain (LC) and a C-terminal heavy chain (HC) which are linked by a disulphide bond. The HC is additionally separated into an N-terminal translocation domain (H_N) and a C-terminal receptor-binding domain (H_C). Further analysis of the novel subtype E1 variants amino acid sequence revealed non-conservative amino acid replacements in the translocation domain (Figure S1). Whether these changes affect neurotoxin translocation into the host cytosol remains unknown.

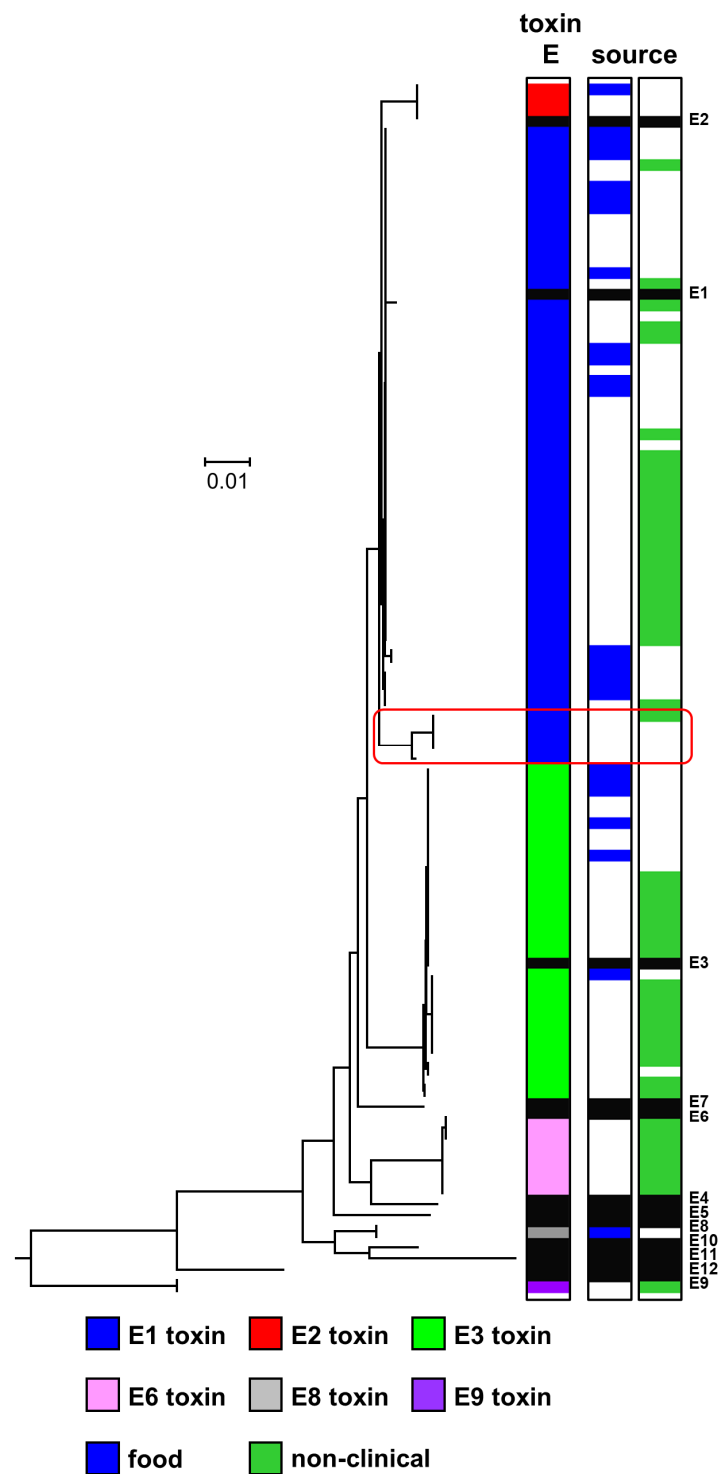


Figure 3. Phylogeny of subtypes of botulinum neurotoxin type E. Further details about individual sequences and isolates are given in Table S2, where they appear in the same order as in this figure. The sources of the isolates are given in Table S1. The toxin protein sequences were aligned with the Muscle module of MEGA7 [59] algorithm and the phylogenetic tree was generated using the Neighbour-Joining method. The new variants of subtype E1 are highlighted in a red box. Scale bar represents the number of amino acid substitutions per site. A total of 112 amino acid sequences were analysed. White blocks represent absence of information regarding source. Black blocks represent reference neurotoxins [16].

Three strains possessed an identical gene encoding subtype E2 botulinum neurotoxin (Figure 3). Two of these strains were isolated following foodborne botulism outbreaks in Alaska [42,45]. The source of the third strain is not known. Thirty strains possessed a gene encoding subtype E3 botulinum neurotoxin (Figure 3). The strains were from diverse geographical locations, including various countries over three continents (Europe, north America and Asia), and were associated with foodborne botulism or isolated from the environment (non-clinical isolates) or from an unknown source (Table S1). Four variants of the subtype E3 botulinum neurotoxin were identified (Figure 3). MacDonald et al. [42] previously found three subtype E3 botulinum neurotoxin variants amongst 16 strains, while Weedmark et al. [60] identified nine subtype E3 variants within 125 subtype E3 strains isolated from widely distributed sites throughout much of northern Canada. The same subtype E3 variant (present in strain Alaska E43) was dominant in both the present and previous studies [42,60]. The seven strains that possessed a gene encoding subtype E6 botulinum neurotoxin were all non-clinical isolates from fresh fish, fish product, or fishpond sediments from different geographical locations in Finland (Table S1). Two variants of the subtype E6 botulinum neurotoxin were identified (Figure 3). Interestingly, none of 175 *C. botulinum* Group II type E strains associated with foodborne botulism or isolated from the environment (non-clinical isolates) in northern Canada possessed a subtype E6 botulinum neurotoxin gene [60].

Botulinum neurotoxin type F is highly diverse. Eight subtypes (F1 to F8) are recognised that differ by up to 36.2% of amino acid residues [15]. Botulinum neurotoxin subtype F6 is exclusively formed by strains of *C. botulinum* Group II, and small differences (amino acid difference of 0.2%) were described between seven examples of this botulinum neurotoxin subtype [3,16]. The 14 strains of *C. botulinum* Group II type F examined in the present study comprised one isolate from an outbreak of foodborne botulism involving venison jerky in USA [61], with most of the remaining strains being non-clinical isolates [25], and one strain unknown. A majority of the strains originated from north America (Table S1). All strains possessed a gene encoding subtype F6 botulinum neurotoxin (Figure 4). Previous studies have reported the subtype F6 neurotoxin gene to be the only type F neurotoxin gene present in *C. botulinum* Group II [3,16]. Two variants of the subtype F6 neurotoxin gene were identified, with the encoded neurotoxins >99% identical. All subtype F6 genes were located within a typical *orfX* neurotoxin gene cluster. Remnants of type B and type E botulinum neurotoxin genes were also detected in all *C. botulinum* Group II type F strains, as described previously [25].

2.2. Whole-Genome Analysis Based on Single-Nucleotide Polymorphisms (SNPs)

C. botulinum Group II is reported to be a more diverse bacterium than *C. botulinum* Group I, with less horizontal gene transfer and fewer recombination events [1,3,6,8,48,49]. Various techniques including comparison of core genome single-nucleotide polymorphisms (SNPs) following whole-genome sequencing and comparative genomic indexing (using a DNA microarray) have been employed to understand the diversity and population structure, and the evolutionary and phylogenetic relationship between strains of *C. botulinum* Group II [2,5,6,8,42–46,62]. In the present study, the genomic diversity of 208 strains of *C. botulinum* Group II has been investigated based on an analysis of core genome SNPs following whole-genome sequencing (Figure 5). Two major lineages have been identified. The first lineage contains strains possessing a type E neurotoxin gene (referred to as type E toxin gene lineage) and is distantly separated from the other lineage. The other lineage (referred to as type B/E/F toxin gene lineage) separates into two closely related sub-lineages, dominated by strains possessing a subtype B4 toxin gene, and also includes strains possessing a subtype F6 toxin gene and a small number of strains with a type E toxin gene (Figure 5). Strains lacking a botulinum neurotoxin gene were present throughout both lineages (Figure 5) and closely related to strains possessing a neurotoxin gene on a plasmid or the chromosome [3,23–28]. Type E and subtype F6 genes were located within an *orfX* neurotoxin gene cluster, while subtype B4 genes were located within a *ha* neurotoxin gene cluster (Figure 5). It is notable that the population and diversity structure identified in the present study contains no additional major lineages, and is consistent with that made in previous studies that

have largely been carried out on a smaller number of and/or less diverse strains [2,5,6,8,42–46,49,62]. This suggests that further strains of *C. botulinum* Group II may keep to the same arrangement.

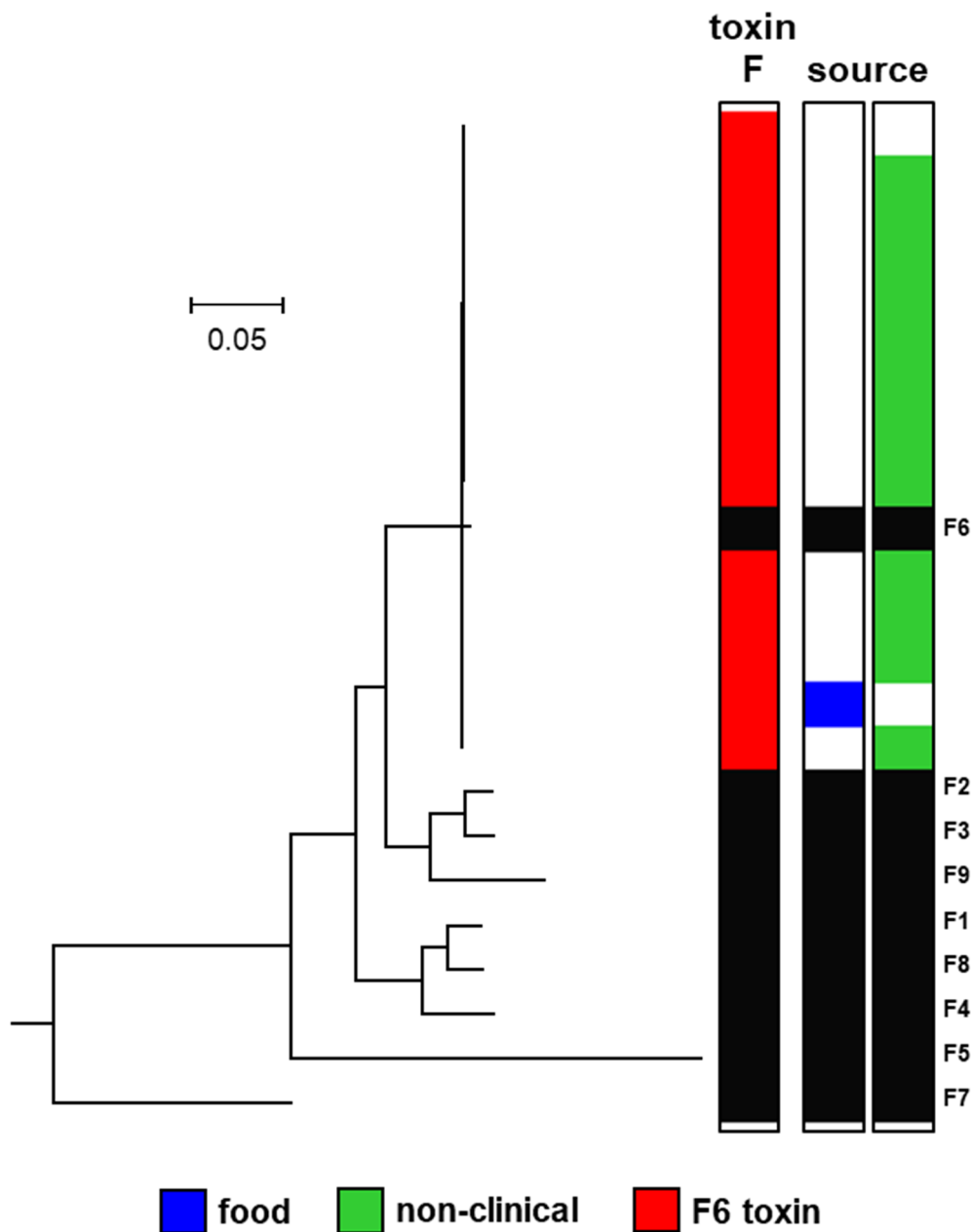


Figure 4. Phylogeny of subtypes of botulinum neurotoxin type F. Further details about individual sequences and isolates are given in Table S2, where they appear in the same order as in this figure. The sources of the isolates are specified in Table S1. The toxin protein sequences were aligned with the Muscle module of MEGA7 [59] algorithm and the phylogenetic tree was generated using the Neighbour-Joining method. Scale bar represents the number of amino acid substitutions per site. A total of 23 amino acid sequences were analysed. White blocks represent absence of information regarding source. Black blocks represent reference neurotoxins [16].

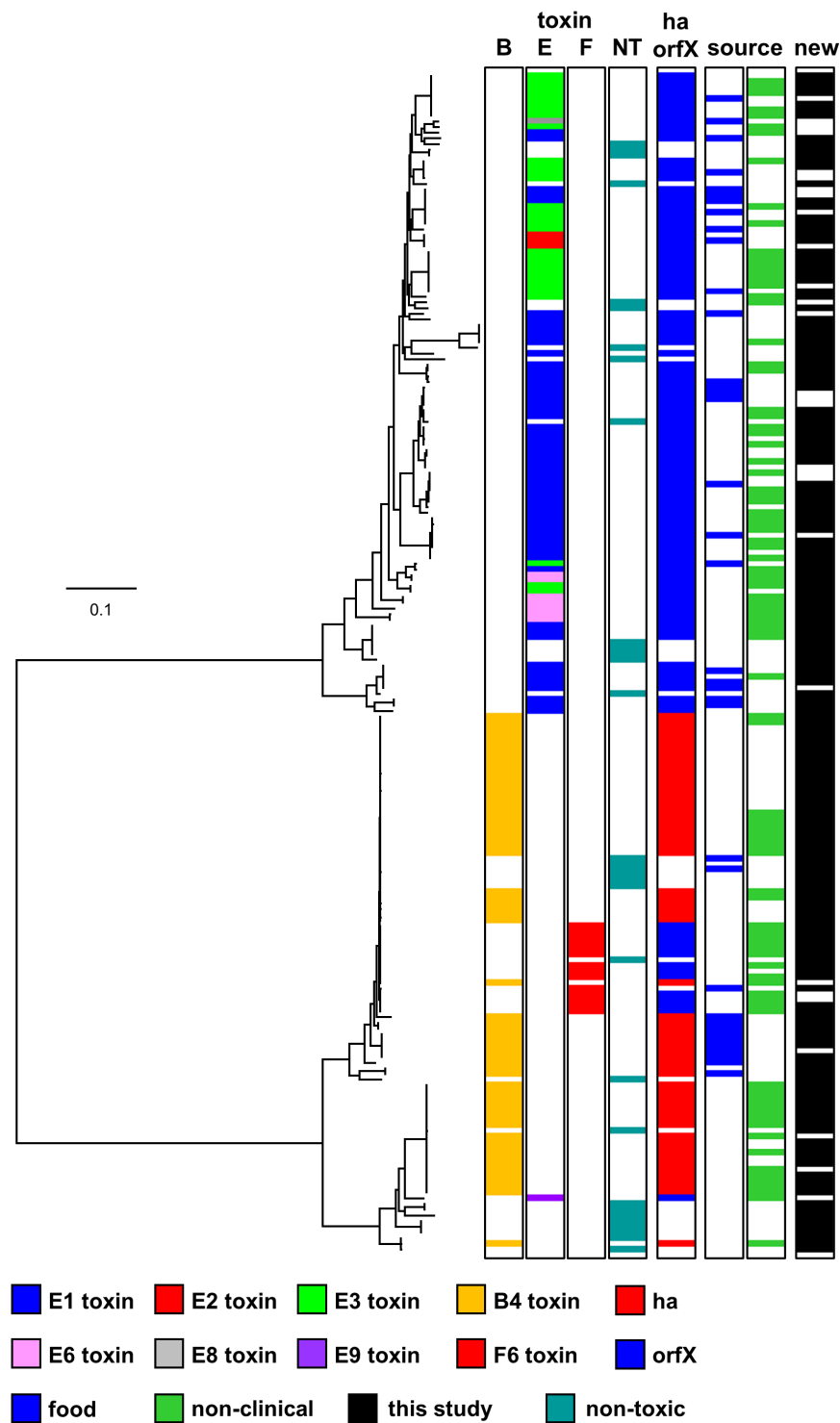


Figure 5. Phylogeny of *C. botulinum* Group II genomes. Two major lineages (type E neurotoxin gene lineage and type B/E/F neurotoxin gene lineage) were identified. The phylogenetic tree was created by comparison of core single-nucleotide polymorphisms identified using the ParSNP program [63]. Treegraph v2 [64], MEGA7 [59] and Figtree were used to annotate and visualize the phylogenetic tree. Accessory genes were identified using BLAST, using reference sequences for the *ha* and *orfX* neurotoxin cluster configurations. The distance bar (0.1) represents the number of nucleotide substitutions per site for a given branch, based on the number of SNPs found in the core genome. Further details about individual sequences and strains are given in Tables S1 and S3, where they appear in the same order as in this figure.

A comparison of the physiological response of strains within these two lineages has revealed a relationship between lineage (and toxin serotype) and carbohydrate utilisation pattern, but not growth response at chill temperature or at high NaCl concentrations [45]. However, a systematic assessment of literature data has established a relationship between lineage (and toxin serotype) and spore heat resistance, despite methodological differences amongst the many studies reviewed [65]. A detailed analysis of the difference in physiological response of strains in these lineages, as well as a genome-wide association study to identify the genetic mechanisms behind each physiological response, is merited. Such information could make a valuable contribution to future microbiological risk assessments [44]. In challenge test experiments where spores from a mixture of strains of *C. botulinum* Group II are added to a food and the potential for neurotoxin formation assessed, it would seem prudent to include strains from both lineages.

The upper lineage shown in Figure 5 includes the majority of strains with genes encoding type E neurotoxin—notably, strains with genes encoding botulinum neurotoxin subtypes E1, E2, E3, E6 and E8, and closely related strains lacking a neurotoxin gene. No strain possessed additional neurotoxin-encoding genes. This is consistent with previous studies [42,44–46]. It has been proposed that these strains and their type E neurotoxin genes arose from the insertion of the type E neurotoxin gene into genetically conserved bacteria, followed by recombination events [42]. Strains possessing a gene encoding subtype E1 neurotoxin appeared more diverse (based on core genome SNPs analysis), and were isolated over a longer period of time and wider geographical location (northern Europe, central Europe, north America, Africa, and Asia) than strains possessing genes encoding subtypes E2, E3 or E6 neurotoxin (Table S1), although it should be noted that more strains possessing a gene encoding subtype E1 neurotoxin were included in the present study. Strains throughout this lineage were associated with foodborne botulism or isolated from the environment (non-clinical isolates) or from an unknown source (Figure 5). Foodborne botulism was generally associated with the consumption of fish.

The two lower sub-lineages shown in Figure 5 are both dominated by strains possessing a subtype B4 toxin gene and are more closely related to each other than to the upper type E lineage. The uppermost sub-lineage contains strains with a gene encoding subtype B4 or subtype F6 neurotoxin and also closely related strains lacking a neurotoxin gene (Figure 5). Strains with a gene encoding subtype B4 neurotoxin possessed no additional neurotoxin-encoding genes, while strains with a gene encoding subtype F6 neurotoxin possessed remnants of other botulinum neurotoxin genes, as noted previously [25]. The strains with a gene encoding subtype F6 neurotoxin are very closely related and are proposed to have arisen once following a series of events that included the insertion of a 34 kb cassette that includes a gene encoding subtype F6 neurotoxin into an ancestral type B strain, and inactivation of the type B gene [25]. Interestingly, strain DB-2, which was collected from Pacific Ocean sediments [23], possessed a subtype B4 neurotoxin gene but is very closely related to strains with a gene encoding subtype F6 neurotoxin (Figure 5). This confirms previous observations based on whole-genome analysis that strain DB-2 is more closely related to *C. botulinum* Group II type F strains 202F and 610F than to other *C. botulinum* Group II type B strains [43,46]. Strains throughout this sub-lineage were associated with foodborne botulism or isolated from the environment (non-clinical isolates) or an unknown source. Foodborne botulism was associated with the consumption of meat or fish contaminated with botulinum neurotoxin in Europe or north America. Although not included in the present study, it is likely (based on whole-genome analysis) that a subtype E12 strain (84–10) isolated in France also locates to this sub-lineage [46]. Thus, this sub-lineage is likely to include *C. botulinum* Group II type B, E, and F strains.

The lower sub-lineage shown in Figure 5 contains strains possessing a subtype B4 toxin gene, a type E toxin gene, and also closely related strains lacking a neurotoxin gene. Strains possessed no more than one neurotoxin-encoding gene. Previous reports indicate that only a small number of strains possessing a type E toxin gene are located in this sub-lineage [27,43,44,46,50,66]. Two of the isolates present within the lower sub-lineage (Figure 5) and lacking a botulinum neurotoxin gene were derived

in our laboratory from strains that possessed a subtype B4 neurotoxin gene. One of these isolates was characterized in some detail. Except for an inability to form botulinum neurotoxin, no physiological differences were identified between the non-toxic isolate (BL81/22NT, CB0395) and the original toxic strain (Eklund 17B(3), CB0597). However, the toxic strain possessed the entire neurotoxin-encoding plasmid p17BNRP, which was completely absent from the non-toxic isolate. None of the strains examined in the present study in the lower sub-lineage were associated with foodborne botulism.

It is important to differentiate *C. botulinum* Groups and indeed their lineages, especially when investigating outbreaks of botulism involving humans or animals. Traditionally, identification relied on growing cultures that would be subsequently tested using either mouse bioassays, ELISAs, mass spectrometry, toxin activity assays, and/or PCR assays for toxin gene presence. One major advantage of the PCR assay is the speed of the test. Williamson and colleagues [47] have developed an in silico multiplex PCR assay to differentiate two major distinct lineages of *C. botulinum* Group II (i.e., type E toxin gene lineage, and type B/E/F toxin gene lineage; Figure 5). The sequencing work in the present study has greatly increased the number of genomes now publicly available, and we have evaluated whether these new additional genomes were correctly assigned using an established in silico multiplex PCR [47]. Of the 208 strains tested in the present study, the in silico PCR assigned 207 strains into the correct major lineage of *C. botulinum* Group II. Only strain IFR 18/139 (CB0543) was not correctly assigned. Upon further analysis, the in silico PCR failed to identify this strain partially due to the target sequence being split over two contigs and additional rare mutations. In reality, conventional PCR would feasibly detect strain IFR 18/139 as in silico PCR is more stringent and requires a complete sequence match. The findings therefore support the continued use of this in silico multiplex PCR [47].

2.3. Pan-genome Comparison of the Two *C. botulinum* Group II Genomic Lineages

The subdivision of the *C. botulinum* Group II genomes into two distinct genomic lineages, one containing 113 genomes with different subtypes of the type E toxin gene (type E toxin gene lineage), and one containing 95 genomes primarily with subtype B4 toxin gene or subtype F6 toxin gene or more rarely type E toxin gene (type B/E/F toxin gene lineage) (Figure 5) was used to search for lineage-specific genes. The pangenome of *C. botulinum* Group II was generated using Prokka-annotated genome sequences and the Roary software package [67], followed by identification of lineage-specific genes defined as present in $\geq 90\%$ of genomes of the lineage, and $\leq 10\%$ of genomes in the other lineage. The *C. botulinum* Group II pan-genome contained a total of 16,571 annotated features (Figure 6A), with a core genome of 1768 features present in $>99\%$ of all genomes from both lineages. There were 390 features specific for the type E toxin gene lineage, and 409 features specific for the type B/E/F toxin gene lineage (Figure 6B). Two complete genomes were selected as representatives for each lineage, Alaska E43 for the type E toxin gene lineage, and Eklund 17B-NRP for the type B/E/F toxin gene lineage, and the relative position of genes specific for their lineage were plotted (Figure 6C). The lineage-specific genes were spread throughout the chromosome, without a clear hotspot for integration such as mobile elements, plasmids or phages. While a detailed investigation of the lineage-specific genes falls outside the scope of this study, we note that the type E toxin gene lineage contains a histidine biosynthetic pathway (*hisSZGDCBHAFIE*) absent from the type B/E/F toxin gene lineage, indicating that histidine metabolism or potential auxotrophy may be used to distinguish these two lineages. Conversely, no comparable features were identified in the type B/E/F toxin gene lineage, and hence this may require more in-depth investigation.

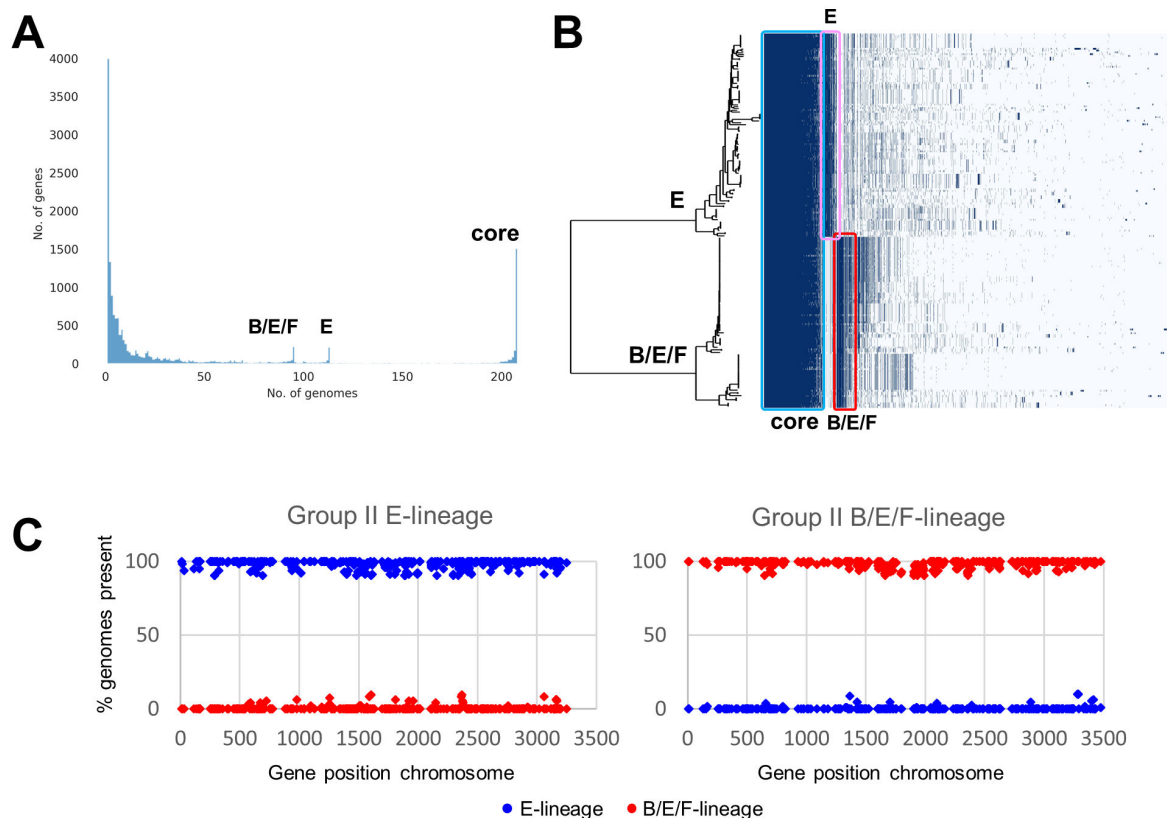


Figure 6. Pan-genome comparison of the two major lineages of *C. botulinum* Group II. The two major lineages (type E neurotoxin gene lineage and type B/E/F neurotoxin gene lineage) are shown in Figure 5. (A) Distribution of features per genome, with many features only present in a few genomes. The core features are shown at the right-hand side, and the type E toxin gene lineage and type B/E/F toxin gene lineage-specific genes are found in the centre. (B) Graphical representation of feature distribution linked to the SNP-based phylogenetic tree also used in Figure 5. The features marked by the blue square represent the core genome, the pink square represents the genes specific for the type E toxin gene lineage, and the red square represents the genes specific for the type B/E/F toxin gene lineage. (C) The lineage-specific genes are distributed throughout the genome, as shown using representative genomes for the type E toxin gene lineage (Alaska E43) and type B/E/F toxin gene lineage (Eklund 17B-NRP). Dark blue diamonds represent type E toxin gene lineage features, and red diamonds represent type B/E/F toxin lineage features.

3. Conclusions

This comparative genomic study of 208 highly diverse strains of *C. botulinum* Group II (180 newly sequenced strains, 28 sequences from Genbank) has considerably increased the number of available genomes for this bacterium. New findings have been made with respect to population structure, diversity and spread, and novel relationships have been established between whole-genome lineage, botulinum neurotoxin subtype variant, association with foodborne botulism, epidemiology, and geographical source. Strains possessed a single type B, E, or F neurotoxin gene. No genes were identified that might encode previously undescribed botulinum neurotoxin serotypes or subtypes, potentially indicating that there are few presently undescribed neurotoxin serotypes or subtypes associated with *C. botulinum* Group II. However, novel botulinum neurotoxin subtype variants were identified, each with a unique amino acid sequence. Core genome SNPs analysis revealed two major lineages—one with type E strains (type E toxin gene lineage), and the second with two closely related sub-lineages dominated by subtype B4 strains that also included subtype F6 strains and a small number of type E strains (type B/E/F toxin gene lineage). Both lineages contained closely related strains that lacked a botulinum neurotoxin-encoding gene. The absence of new lineages and the finding that

the population diversity organisation is similar to that described previously based on fewer and/or less diverse strains suggest that further strains of *C. botulinum* Group II may also follow the same organisation. The *C. botulinum* Group II pan-genome contained 16,571 annotated features, with a core genome of 1768 features present in >99% of all genomes, 390 features specific to the type E toxin gene lineage, and 490 features specific to the type B/E/F toxin gene lineage. The lineage-specific genes were spread throughout the chromosome, with no integration hotspot. Finally, the genome sequences are a valuable resource—for example, contributing to future research on the evolution of *C. botulinum* Group II and its neurotoxin genes, the dissection of key aspects of the biology of *C. botulinum* Group II essential for food chain transmission, and the targeting of PCR assays to detect and distinguish strains. This may contribute to improved microbiological risk assessments, the safe development of novel “fresh-like” foods and the prevention of foodborne botulism.

4. Materials and Methods

4.1. *C. botulinum* Group II Isolates

The genomes of 208 isolates were assigned to *C. botulinum* Group II based on their genome sequence. This comprised 180 newly sequenced genomes, and 28 genomes from public sources. The newly sequenced genomes either possessed a single neurotoxin gene of subtype B4 (63 strains), E1 (58 strains), E2 (3 strains), E3 (30 strains), E6 (7 strains), F6 (14 strains) or no neurotoxin gene (33 strains). Strains with a complete subtype F6 gene also contained fragments of other neurotoxin genes. It was not determined whether the presence of a neurotoxin gene was associated with formation of botulinum neurotoxin. The newly sequenced strains were from 16 countries and four continents (Figure 1); with the first strain isolated in 1934. Thirty-seven strains were associated with outbreaks of foodborne botulism, 94 strains were non-clinical isolates not associated with foodborne botulism outbreaks (i.e., isolated from the environment), and the status of the remaining strains is unknown (Table S1).

4.2. Genomic DNA Preparation

DNA extraction was performed by following the traditional Gram-positive cell lysis procedures. Briefly, cell pellets from an overnight 10 mL TPYG broth culture (30 °C) were treated with lysozyme, mutanolysin and RNase A in TE buffer at 37 °C, 60 min, followed by lysis in proteinase K, SDS and NaCl for a further 60 min at 55 °C. The clear lysate was purified by phenol-chloroform, then chloroform/IAA extraction before ethanol precipitation and dissolution in TE buffer [68].

4.3. Whole-Genome Sequencing

Whole-genome sequencing of the *C. botulinum* gDNA samples was carried out at the Earlham Institute, Norwich, UK. Libraries were constructed using 1 ng of input DNA in their low input transposase-enabled (LITE) low-cost, high-throughput library construction pipeline based on the Illumina Nextera kits (Illumina, Cambridge, UK). Each library was constructed using unique 9 bp dual index combinations allowing samples to be multiplexed. Pooled libraries were then sequenced with a 2 × 250 bp read metric on an Illumina HiSeq 2500 sequencer.

4.4. Genome Assembly and Quality Control

Genomes were assembled using Shovill version 1.0.4 (<https://github.com/tseemann/shovill>), using the Spades 3.13.0 assembler module [69] and Trimmomatic [70] enabled. Contigs < 200 bp and coverage < 5-fold were excluded from further analyses. Quality checks of the assemblies were performed using QUAST version 4.6 [71]. Genome assemblies were annotated using Prokka version 1.13 [72]. Individual Genbank/SRA accession numbers for FASTQ reads, genome assemblies, and source are given in Table S1.

4.5. Identification of Botulinum Neurotoxin Subtypes and Accessory Protein Configuration

The Prokka-annotated features were searched with representatives of the A, B, C, D, E, F, G, FA, X and Ebont/J toxins [9,16] using BLAST in the BioEdit version 7.2.5 program [73]). Incomplete, fragmented and inactivated toxins were manually generated from the individual genome sequences. Toxins were assigned to specific types after alignment of the toxin amino acid sequences with the reference types [16] using the Muscle module of MEGA7 [59], and subsequent generation of a Neighbour-Joining tree and clustering with reference toxins. Accessory genes were identified using BLAST, using reference sequences for the *ha* and *orfX* configurations [3].

4.6. Pangenome Analyses, Target Gene Identification and in Silico PCR

The *C. botulinum* Group II pangenome was identified using Roary version 3.12 [67], using the Prokka-annotated assemblies, with 80% and 90% BLAST cut-off percentages, and paralog clustering both switched off and on [74]. The Roary outputs were subsequently analysed using Scoary [75], with the no-pairwise option to speed up analysis, and a Bonferroni corrected *p*-value cut-off of 0.05. Gene distribution was studied and only genes present in $\geq 90\%$ in the tester group and less than $\leq 10\%$ of the comparator group were further studied. The plots for the pangenome distribution were generated with roary_plots (https://github.com/sanger-pathogens/Roary/tree/master/contrib/roary_plots). In silico PCR was performed using MIST [76].

4.7. Core Genome SNPs Analysis and Phylogenetic Trees

Core genome SNPs were identified using the ParSNP program [63] with the “-x -a 13” settings [77]. Treegraph v2 [64], MEGA7 [59] and Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>) were used to annotate and visualize the phylogenetic trees.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6651/12/5/306/s1>, Table S1, Strains and genome sequences of *C. botulinum* Group II used in this study; Table S2, Strain ID Figures 2–4; Table S3, Strain ID Figure 5; Figure S1, Alignment of the reference subtype E1 botulinum toxin.

Author Contributions: M.W.P. conceptualized and secured funding for this study. M.L. provided strains for use in the study. J.B., S.C.S. and A.T.C. cultured the bacterial strains, confirmed their purity and properties, and prepared DNA for whole-genome sequencing. M.W.P. verified the origin of the strains. A.H.M.v.V. carried out the analysis of genomic data and data visualisation, with input from J.B. A.H.M.v.V. submitted genome sequence data to GenBank. J.B. and M.W.P. prepared the original draft manuscript and co-ordinated manuscript finalisation, with all authors reviewing, editing and providing critical feedback. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the BBSRC Institute Strategic Programme on Gut Health and Food Safety [grant number BB/J004529/1] and by the European Research Council [grant number ERC-CoG 683099 ‘whyBOther’]. The authors are very grateful to colleagues (C. Hatheway, A. East, K. Grant, P. Barrett and J. Crowther) for providing additional strains used in the present study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bruggemann, H.; Wollherr, A.; Mazuet, C.; Popoff, M. *Clostridium botulinum*. In *Genomes of Foodborne and Waterborne Pathogens*; Fratamico, P., Liu, Y., Kathariou, S., Eds.; ASM Press: Washington, DC, USA, 2011; pp. 185–212. ISBN 978-1555814571.
- Carter, A.T.; Paul, C.J.; Mason, D.R.; Twine, S.M.; Alston, M.J.; Logan, S.M.; Austin, J.W.; Peck, M.W. Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. *BMC Genom.* **2009**, *10*, 115. [CrossRef] [PubMed]
- Carter, A.T.; Peck, M.W. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. *Res. Microbiol.* **2015**, *166*, 303–317. [CrossRef] [PubMed]
- Johnson, E.A. *Food Microbiology: Fundamentals and Frontiers*; Doyle, M.P., Buchanan, R.L., Eds.; ASM Press: Washington, DC, USA, 2013; pp. 441–464.

5. Hill, K.K.; Smith, T.J.; Helma, C.H.; Ticknor, L.O.; Foley, B.T.; Svensson, R.T.; Brown, J.L.; Johnson, E.A.; Smith, L.A.; Okinaka, R.T.; et al. Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J. Bacteriol.* **2007**, *189*, 818–832. [[CrossRef](#)] [[PubMed](#)]
6. Peck, M.W. Biology and genomic analysis of *Clostridium botulinum*. *Adv. Microb. Physiol.* **2009**, *55*, 183–320. [[CrossRef](#)] [[PubMed](#)]
7. Smith, T.; Williamson, C.H.D.; Hill, K.; Sahl, J.; Keim, P. Botulinum Neurotoxin-Producing Bacteria. Isn't It Time that We Called a Species a Species? *MBio* **2018**, *9*, e01469–18. [[CrossRef](#)]
8. Williamson, C.H.; Sahl, J.W.; Smith, T.J.; Xie, G.; Foley, B.T.; Smith, L.A.; Fernandez, R.A.; Lindström, M.; Korkeala, H.; Keim, P.; et al. Comparative genomic analyses reveal broad diversity in botulinum-toxin-producing Clostridia. *BMC Genom.* **2016**, *17*, 180. [[CrossRef](#)]
9. Brunt, J.; Carter, A.T.; Stringer, S.C.; Peck, M.W. Identification of a novel botulinum neurotoxin gene cluster in Enterococcus. *FEBS Lett.* **2018**, *592*, 310–317. [[CrossRef](#)]
10. Mansfield, M.J.; Doxey, A.C. Genomic insights into the evolution and ecology of botulinum neurotoxins. *Pathog. Dis.* **2018**, *76*. [[CrossRef](#)]
11. Zhang, S.; Lebreton, F.; Mansfield, M.J.; Miyashita, S.I.; Zhang, J.; Schwartzman, J.A.; Tao, L.; Masuyer, G.; Martinez-Carranza, M.; Stenmark, P.; et al. Identification of a Botulinum Neurotoxin-like Toxin in a Commensal Strain of *Enterococcus faecium*. *Cell Host Microbe* **2018**, *23*, 169–176. [[CrossRef](#)]
12. Hatheway, C.L. Toxigenic clostridia. *Clin. Microbiol. Rev.* **1990**, *3*, 66–98. [[CrossRef](#)]
13. Poulain, B.; Popoff, M.R. Why Are Botulinum Neurotoxin-Producing Bacteria So Diverse and Botulinum Neurotoxins So Toxic? *Toxins* **2019**, *11*, 34. [[CrossRef](#)] [[PubMed](#)]
14. Rossetto, O.; Pirazzini, M.; Montecucco, C. Botulinum neurotoxins: Genetic, structural and mechanistic insights. *Nat. Rev. Microbiol.* **2014**, *12*, 535–549. [[CrossRef](#)] [[PubMed](#)]
15. Rummel, A. The long journey of botulinum neurotoxins into the synapse. *Toxicon* **2015**, *107*, 9–24. [[CrossRef](#)] [[PubMed](#)]
16. Peck, M.W.; Smith, T.J.; Anniballi, F.; Austin, J.W.; Bano, L.; Bradshaw, M.; Cuervo, P.; Cheng, L.W.; Derman, Y.; Dorner, B.G.; et al. Historical Perspectives and Guidelines for Botulinum Neurotoxin Subtype Nomenclature. *Toxins* **2017**, *9*, 38. [[CrossRef](#)] [[PubMed](#)]
17. Smith, T.J.; Hill, K.K.; Raphael, B.H. Historical and current perspectives on *Clostridium botulinum* diversity. *Res. Microbiol.* **2015**, *166*, 290–302. [[CrossRef](#)]
18. Kaji, R. Clinical differences between A1 and A2 botulinum toxin subtypes. *Toxicon* **2015**, *107*, 85–88. [[CrossRef](#)]
19. Moritz, M.S.; Tepp, W.H.; Bradshaw, M.; Johnson, E.A.; Pellett, S. Isolation and Characterization of the Novel Botulinum Neurotoxin A Subtype 6. *mSphere* **2018**, *3*. [[CrossRef](#)]
20. Moritz, M.S.; Tepp, W.H.; Inzalaco, H.N.T.; Johnson, E.A.; Pellett, S. Comparative functional analysis of mice after local injection with botulinum neurotoxin A1, A2, A6, and B1 by catwalk analysis. *Toxicon* **2019**, *167*, 20–28. [[CrossRef](#)]
21. Pellett, S.; Tepp, W.H.; Whitmarsh, R.C.; Bradshaw, M.; Johnson, E.A. In vivo onset and duration of action varies for botulinum neurotoxin a subtypes 1–5. *Toxicon* **2015**, *107*, 37–42. [[CrossRef](#)]
22. Whitmarsh, R.C.; Tepp, W.H.; Bradshaw, M.; Lin, G.; Pier, C.L.; Scherf, J.M.; Johnson, E.A.; Pellett, S. Characterization of botulinum neurotoxin A subtypes 1 through 5 by investigation of activities in mice, in neuronal cell cultures, and in vitro. *Infect. Immun.* **2013**, *81*, 3894–3902. [[CrossRef](#)]
23. Carter, A.T.; Austin, J.W.; Weedmark, K.A.; Corbett, C.; Peck, M.W. Three classes of plasmid (47–63 kb) carry the type B neurotoxin gene cluster of group II *Clostridium botulinum*. *Genome Biol. Evol.* **2014**, *6*, 2076–2087. [[CrossRef](#)] [[PubMed](#)]
24. Carter, A.T.; Austin, J.W.; Weedmark, K.A.; Peck, M.W. Evolution of Chromosomal *Clostridium botulinum* Type E Neurotoxin Gene Clusters: Evidence Provided by Their Rare Plasmid-Borne Counterparts. *Genome Biol. Evol.* **2016**, *8*, 540–555. [[CrossRef](#)] [[PubMed](#)]
25. Carter, A.T.; Stringer, S.C.; Webb, M.D.; Peck, M.W. The type F6 neurotoxin gene cluster locus of group II *Clostridium botulinum* has evolved by successive disruption of two different ancestral precursors. *Genome Biol. Evol.* **2013**, *5*, 1032–1037. [[CrossRef](#)]
26. Hill, K.K.; Smith, T.J. Genetic diversity within *Clostridium botulinum* serotypes, botulinum neurotoxin gene clusters and toxin subtypes. *Curr. Top. Microbiol. Immunol.* **2013**, *364*, 1–20. [[CrossRef](#)]
27. Hill, K.K.; Xie, G.; Foley, B.T.; Smith, T.J. Genetic diversity within the botulinum neurotoxin-producing bacteria and their neurotoxins. *Toxicon* **2015**, *107*, 2–8. [[CrossRef](#)]

28. Zhang, Z.; Hintsala, H.; Chen, Y.; Korkeala, H.; Lindström, M. Plasmid-Borne Type E Neurotoxin Gene Clusters in *Clostridium botulinum* Strains. *Appl. Environ. Microbiol.* **2013**, *79*, 3856–3859. [[CrossRef](#)]
29. Chen, Y.; Korkeala, H.; Aarnikunnas, J.; Lindström, M. Sequencing the botulinum neurotoxin gene and related genes in *Clostridium botulinum* type E strains reveals orfX3 and a novel type E neurotoxin subtype. *J. Bacteriol.* **2007**, *189*, 8643–8650. [[CrossRef](#)]
30. Kubota, T.; Yonekura, N.; Hariya, Y.; Isogai, E.; Isogai, H.; Amano, K.; Fujii, N. Gene arrangement in the upstream region of *Clostridium botulinum* type E and *Clostridium butyricum* BL6340 progenitor toxin genes is different from that of other types. *FEMS Microbiol. Lett.* **1998**, *158*, 215–221. [[CrossRef](#)]
31. Nowakowska, M.B.; Douillard, F.P.; Lindström, M. Looking for the X Factor in Bacterial Pathogenesis: Association of orfX-p47 Gene Clusters with Toxin Genes in Clostridial and Non-Clostridial Bacterial Species. *Toxins* **2019**, *12*, 19. [[CrossRef](#)]
32. Barker, G.C.; Malakar, P.K.; Plowman, J.; Peck, M.W. Quantification of Nonproteolytic *Clostridium botulinum* Spore Loads in Food Materials. *Appl. Environ. Microbiol.* **2016**, *82*, 1675–1685. [[CrossRef](#)]
33. Hielm, S.; Björkroth, J.; Hyytiä, E.; Korkeala, H. Prevalence of *Clostridium botulinum* in Finnish trout farms: Pulsed-field gel electrophoresis typing reveals extensive genetic diversity among type E isolates. *Appl. Environ. Microbiol.* **1998**, *64*, 4161–4167. [[CrossRef](#)] [[PubMed](#)]
34. Hyytiä, E.; Hielm, S.; Korkeala, H. Prevalence of *Clostridium botulinum* type E in Finnish fish and fishery products. *Epidemiol. Infect.* **1998**, *120*, 245–250. [[CrossRef](#)] [[PubMed](#)]
35. Peck, M.W.; Plowman, J.; Aldus, C.F.; Wyatt, G.M.; Penalzoza Izurieta, W.; Stringer, S.C.; Barker, G.C. Development and Application of a New Method for Specific and Sensitive Enumeration of Spores of Nonproteolytic *Clostridium botulinum* Types B, E, and F in Foods and Food Materials. *Appl. Environ. Microbiol.* **2010**, *76*, 6607. [[CrossRef](#)] [[PubMed](#)]
36. Smith, L.D.S.; Sugiyama, H. *Botulism: The Organism, Its Toxins, the Disease*, 2nd ed.; Charles C Thomas: Springfield, IL, USA, 1988; p. 171.
37. Van Ermengem, E. Classics in infectious diseases. A new anaerobic bacillus and its relation to botulism. E. van Ermengem. Originally published as “Ueber einen neuen anaeroben Bacillus und seine Beziehungen zum Botulismus” in Zeitschrift für Hygiene und Infektionskrankheiten 26: 1–56, 1897. *Rev. Infect. Dis.* **1979**, *1*, 701–719.
38. Peck, M.W. *Clostridium botulinum* and the safety of minimally heated, chilled foods: An emerging issue? *J. Appl. Microbiol.* **2006**, *101*, 556–570. [[CrossRef](#)]
39. Peck, M.W.; Goodburn, K.E.; Betts, R.P.; Stringer, S.C. Assessment of the potential for growth and neurotoxin formation by non-proteolytic *Clostridium botulinum* in short shelf-life commercial foods designed to be stored chilled. *Trends Food Sci. Tech.* **2008**, *19*, 207–216. [[CrossRef](#)]
40. Peck, M.W.; Stringer, S.C.; Carter, A.T. *Clostridium botulinum* in the post-genomic era. *Food Microbiol.* **2011**, *28*, 183–191. [[CrossRef](#)]
41. Peck, M.W.; Webb, M.D.; Goodburn, K.E. Assessment of the risk of botulism from chilled, vacuum/modified atmosphere packed fresh beef, lamb and pork held at 3 °C to 8 °C. *Food Microbiol.* **2020**, in press.
42. Macdonald, T.E.; Helma, C.H.; Shou, Y.; Valdez, Y.E.; Ticknor, L.O.; Foley, B.T.; Davis, S.W.; Hannett, G.E.; Kelly-Cirino, C.D.; Barash, J.R.; et al. Analysis of *Clostridium botulinum* serotype E strains by using multilocus sequence typing, amplified fragment length polymorphism, variable-number tandem-repeat analysis, and botulinum neurotoxin gene sequencing. *Appl. Environ. Microbiol.* **2011**, *77*, 8625–8634. [[CrossRef](#)]
43. Mazuet, C.; Legeay, C.; Sautereau, J.; Ma, L.; Bouchier, C.; Bouvet, P.; Popoff, M.R. Diversity of Group I and II *Clostridium botulinum* Strains from France Including Recently Identified Subtypes. *Genome Biol. Evol.* **2016**, *8*, 1643–1660. [[CrossRef](#)]
44. Peck, M.W.; van Vliet, A.H. Impact of *Clostridium botulinum* genomic diversity on food safety. *Curr. Opin. Food Sci.* **2016**, *10*, 52–59. [[CrossRef](#)] [[PubMed](#)]
45. Stringer, S.C.; Carter, A.T.; Webb, M.D.; Wachnicka, E.; Crossman, L.C.; Sebaihia, M.; Peck, M.W. Genomic and physiological variability within Group II (non-proteolytic) *Clostridium botulinum*. *BMC Genom.* **2013**, *14*, 333. [[CrossRef](#)] [[PubMed](#)]
46. Weedmark, K.A.; Mabon, P.; Hayden, K.L.; Lambert, D.; Van Domselaar, G.; Austin, J.W.; Corbett, C.R. *Clostridium botulinum* Group II Isolate Phylogenomic Profiling Using Whole-Genome Sequence Data. *Appl. Environ. Microbiol.* **2015**, *81*, 5938–5948. [[CrossRef](#)]

47. Williamson, C.H.D.; Vazquez, A.J.; Hill, K.; Smith, T.J.; Nottingham, R.; Stone, N.E.; Sobek, C.J.; Cocking, J.H.; Fernandez, R.A.; Caballero, P.A.; et al. Differentiating Botulinum Neurotoxin-Producing Clostridia with a Simple, Multiplex PCR Assay. *Appl. Environ. Microbiol.* **2017**, *83*, e00806–e00817. [\[CrossRef\]](#)
48. Keto-Timonen, R.; Heikinheimo, A.; Eerola, E.; Korkeala, H. Identification of Clostridium species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. *J. Clin. Microbiol.* **2006**, *44*, 4057–4065. [\[CrossRef\]](#)
49. Keto-Timonen, R.; Nevas, M.; Korkeala, H. Efficient DNA fingerprinting of *Clostridium botulinum* types A, B, E, and F by amplified fragment length polymorphism analysis. *Appl. Environ. Microbiol.* **2005**, *71*, 1148–1154. [\[CrossRef\]](#)
50. Raphael, B.H.; Lautenschlager, M.; Kalb, S.R.; de Jong, L.I.; Frace, M.; Luquez, C.; Barr, J.R.; Fernandez, R.A.; Maslanka, S.E. Analysis of a unique *Clostridium botulinum* strain from the Southern hemisphere producing a novel type E botulinum neurotoxin subtype. *BMC Microbiol.* **2012**, *12*, 245. [\[CrossRef\]](#)
51. Rasetti-Escargueil, C.; Lemichez, E.; Popoff, M.R. Public Health Risk Associated with Botulism as Foodborne Zoonoses. *Toxins* **2019**, *12*, 17. [\[CrossRef\]](#)
52. Brunt, J.; Carter, A.T.; Pye, H.V.; Peck, M.W. The orphan germinant receptor protein GerXAO (but not GerX3b) is essential for L-alanine induced germination in *Clostridium botulinum* Group II. *Sci. Rep.* **2018**, *8*, 7060. [\[CrossRef\]](#)
53. Brunt, J.; van Vliet, A.H.M.; van den Bos, F.; Carter, A.T.; Peck, M.W. Diversity of the Germination Apparatus in *Clostridium botulinum* Groups I, II, III, and IV. *Front. Microbiol.* **2016**, *7*, 1702. [\[CrossRef\]](#)
54. Söderholm, H.; Jaakkola, K.; Somervuo, P.; Laine, P.; Auvinen, P.; Paulin, L.; Lindström, M.; Korkeala, H. Comparison of *Clostridium botulinum* genomes shows the absence of cold shock protein coding genes in type E neurotoxin producing strains. *Botulinum J.* **2013**, *2*. [\[CrossRef\]](#)
55. Artin, I.; Carter, A.T.; Holst, E.; Lövenklev, M.; Mason, D.R.; Peck, M.W.; Rådström, P. Effects of Carbon Dioxide on Neurotoxin Gene Expression in Nonproteolytic *Clostridium botulinum* Type E. *Appl. Environ. Microbiol.* **2008**, *74*, 2391. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Mascher, G.; Mertaola, A.; Korkeala, H.; Lindström, M. Neurotoxin synthesis is positively regulated by the sporulation transcription factor Spo0A in *Clostridium botulinum* type E. *Environ. Microbiol.* **2017**, *19*, 4287–4300. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Chen, Y.; Korkeala, H.; Linden, J.; Lindström, M. Quantitative real-time reverse transcription-PCR analysis reveals stable and prolonged neurotoxin cluster gene activity in a *Clostridium botulinum* type E strain at refrigeration temperature. *Appl. Environ. Microbiol.* **2008**, *74*, 6132–6137. [\[CrossRef\]](#)
58. McLauchlin, J.; Grant, K.A.; Little, C.L. Food-borne botulism in the United Kingdom. *J. Public Health* **2006**, *28*, 337–342. [\[CrossRef\]](#)
59. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [\[CrossRef\]](#)
60. Weedmark, K.A.; Lambert, D.L.; Mabon, P.; Hayden, K.L.; Urfano, C.J.; Leclair, D.; Van Domselaar, G.; Austin, J.W.; Corbett, C.R. Two novel toxin variants revealed by whole-genome sequencing of 175 *Clostridium botulinum* type E strains. *Appl. Environ. Microbiol.* **2014**, *80*, 6334–6345. [\[CrossRef\]](#)
61. Midura, T.F.; Nygaard, G.S.; Wood, R.M.; Bodily, H.L. *Clostridium botulinum* type F: Isolation from venison jerky. *Appl. Microbiol.* **1972**, *24*, 165–167. [\[CrossRef\]](#)
62. Hill, K.K.; Xie, G.; Foley, B.T.; Smith, T.J.; Munk, A.C.; Bruce, D.; Smith, L.A.; Brettin, T.S.; Detter, J.C. Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. *BMC Biol.* **2009**, *7*, 66. [\[CrossRef\]](#)
63. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **2014**, *15*, 524. [\[CrossRef\]](#)
64. Stover, B.C.; Muller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* **2010**, *11*, 7. [\[CrossRef\]](#)
65. Wachnicka, E.; Stringer, S.C.; Barker, G.C.; Peck, M.W. Systematic Assessment of Nonproteolytic *Clostridium botulinum* Spores for Heat Resistance. *Appl. Environ. Microbiol.* **2016**, *82*, 6019–6029. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Mazuet, C.; Sautereau, J.; Legeay, C.; Bouchier, C.; Bouvet, P.; Popoff, M.R. An atypical outbreak of food-borne botulism due to *Clostridium botulinum* types B and E from ham. *J. Clin. Microbiol.* **2015**, *53*, 722–726. [\[CrossRef\]](#)

67. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
68. Sebaihia, M.; Peck, M.W.; Minton, N.P.; Thomson, N.R.; Holden, M.T.; Mitchell, W.J.; Carter, A.T.; Bentley, S.D.; Mason, D.R.; Crossman, L.; et al. Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res.* **2007**, *17*, 1082–1092. [[CrossRef](#)]
69. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
70. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
71. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
72. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
73. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **1999**, *41*, 95–98.
74. Van Vliet, A.H.M. Use of pan-genome analysis for the identification of lineage-specific genes of *Helicobacter pylori*. *FEMS Microbiol. Lett.* **2017**, *364*. [[CrossRef](#)]
75. Brynildsrud, O.; Bohlin, J.; Scheffer, L.; Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **2016**, *17*, 238. [[CrossRef](#)] [[PubMed](#)]
76. Kruczkiewicz, P.; Mutschall, S.; Barker, D.; Thomas, J.; Van Domselaar, G.; Gannon, V.P.J.; Carrillo, C.D.; Taboada, E.N. MIST: A Tool for Rapid in silico Generation of Molecular Data from Bacterial Genome Sequences. In Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms, Barcelona, Spain, 11–14 February 2013; pp. 316–323.
77. Pornsukarom, S.; van Vliet, A.H.M.; Thakur, S. Whole genome sequencing analysis of multiple *Salmonella* serovars provides insights into phylogenetic relatedness, antimicrobial resistance, and virulence markers across humans, food animals and agriculture environmental sources. *BMC Genom.* **2018**, *19*, 801. [[CrossRef](#)]

